

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 09-258772

(43)Date of publication of application : 03.10.1997

(51)Int.Cl.

G10L 3/00

G10L 3/00

G10L 3/00

G10L 9/16

(21)Application number : 08-068046

(71)Applicant : CANON INC

(22)Date of filing : 25.03.1996

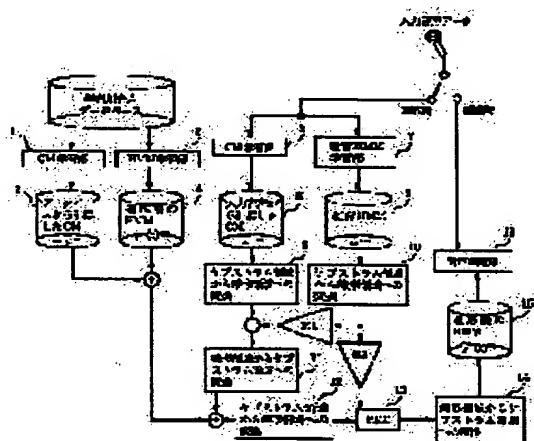
(72)Inventor : KOSAKA TETSUO  
OHORA YASUNORI

## (54) VOICE RECOGNIZING METHOD AND DEVICE

## (57)Abstract:

**PROBLEM TO BE SOLVED:** To obtain a CM with the influence of noise added so as to heighten the accuracy of estimating circuit fluctuation by subtracting a cepstrum long-time average of a voiceless part from a cepstrum long-time average (CCM) of a voice part on a linear spectral dimension.

**SOLUTION:** Learning is performed by a noise HMM (Hidden Markov Model) learning part 7, using the data of a voiceless part of input voice to obtain a noise HMM 8. The long-time average CM 6 of cepstrum of a voice part of the input voice is computed by a CM learning part 5. The CM 6 and the noise HMM 8 are respectively converted from an area into a linear spectral area by converting parts 9, 10. The noise HMM 8 is subtracted from the CM 6, and its result is converted from linear spectral expression into cepstrum expression by a converting part 11. An HMM obtained by a PMC method in a PMC 13 is converted into the cepstrum expression, and voice recognition is performed by a voice recognizing part 16 using an obtained HMM 14.



## LEGAL STATUS

[Date of request for examination] 20.11.2000

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the  
examiner's decision of rejection or application  
converted registration]

[Date of final disposal for application]

[Patent number] 3397568

[Date of registration] 14.02.2003

[Number of appeal against examiner's decision of  
rejection][Date of requesting appeal against examiner's decision  
of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平9-258772

(43) 公開日 平成9年(1997)10月3日

(51) Int.Cl. <sup>6</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G 1 0 L 3/00	5 3 5		G 1 0 L 3/00	5 3 5
	5 2 1			5 2 1 L
	5 6 1			5 6 1 C
9/16	3 0 1		9/16	3 0 1 A

審査請求 未請求 請求項の数18 O L (全 6 頁)

(21) 出願番号 特願平8-68046  
(22) 出願日 平成8年(1996)3月25日

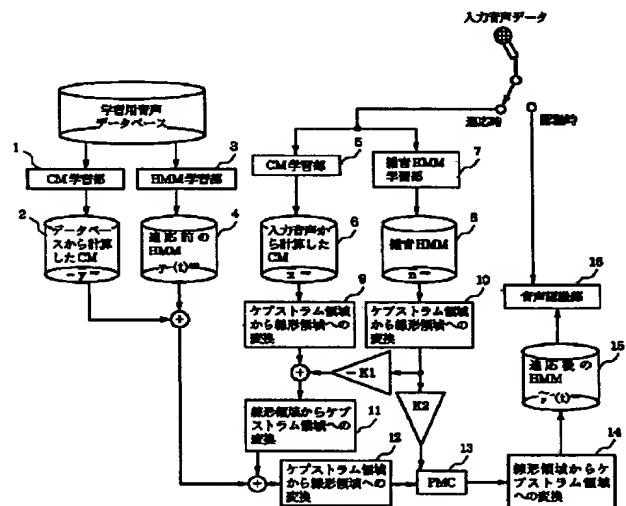
(71) 出願人 000001007  
キヤノン株式会社  
東京都大田区下丸子3丁目30番2号  
(72) 発明者 小坂 哲夫  
東京都大田区下丸子3丁目30番2号キヤノン株式会社内  
(72) 発明者 大洞 恭則  
東京都大田区下丸子3丁目30番2号キヤノン株式会社内  
(74) 代理人 弁理士 丸島 儀一

(54) 【発明の名称】 音声認識方法及び装置

(57) 【要約】

【課題】 音声認識する際に、CMS法とPMS法とを単に組み合わせた場合には、回線特性の推定がうまくいかない。

【解決手段】 音声部と非音声部とを含む音声を入力し、音声入力の音声部から音声部のケプストラム長時間平均を求め（CM学習部）、前記入力音声の非音声部から非音声部のケプストラム長時間平均を求め（雑音HMM学習部）、各ケプストラム長時間平均をケプストラム領域から線形領域に変換した後線形スペクトル次元上で差し引き、このデータに基づいて音声認識を行う。



## 【特許請求の範囲】

【請求項1】 音声部と非音声部とを含む音声を入力し、  
前記入力音声に含まれる音声部から音声部のケプストラム長時間平均を求め、  
前記入力音声に含まれる非音声部から非音声部のケプストラム長時間平均を求め、  
前記音声部のケプストラム長時間平均から前記非音声部のケプストラム長時間平均を線形スペクトル次元上で差し引くことを特徴とする音声認識方法。

【請求項2】 前記差し引いて求めたデータにより入力音声进行を認識することを特徴とする請求項1に記載の音声認識方法。

【請求項3】 前記ケプストラム長時間平均を隠れマルコフモデルのデータとすることを特徴とする請求項1に記載の音声認識方法。

【請求項4】 雑音データと音声データとを線形スペクトル次元で加算合成したデータを用いて雑音重畳モデルを作成し、  
前記雑音重畳モデルを用いて前記非音声部のケプストラム長時間平均を求めることを特徴とする請求項1に記載の音声認識方法。

【請求項5】 前記各処理の制御プログラムを記憶した記憶部を利用し、当該制御プログラムに従って各処理を実行することを特徴とする請求項1に記載の音声認識方法。

【請求項6】 前記音声認識は、隠れマルコフモデルによる音声認識とすることを特徴とする請求項2に記載の音声認識方法。

【請求項7】 前記認識結果を表示部に表示することを特徴とする請求項2に記載の音声認識方法。

【請求項8】 前記認識結果を印字部に印字することを特徴とする請求項2に記載の音声認識方法。

【請求項9】 前記各処理の制御プログラムを記憶した記憶媒体を利用し、当該制御プログラムに従って各処理を実行することを特徴とする請求項1に記載の音声認識方法。

【請求項10】 音声部と非音声部とを含む音声を入力する音声入力手段と、  
前記入力音声に含まれる音声部から音声部のケプストラム長時間平均と、前記入力音声に含まれる非音声部から非音声部のケプストラム長時間平均を求めるケプストラム長時間平均導出手段と、  
前記音声部のケプストラム長時間平均から前記非音声部のケプストラム長時間平均を線形スペクトル次元上で差し引く差し引き手段とを有することを特徴とする音声認識装置。

【請求項11】 前記差し引き手段により差し引いて求めたデータにより音声进行を認識する音声認識手段を有することを特徴とする請求項10に記載の音声認識装置。

【請求項12】 前記ケプストラム長時間平均を隠れマルコフモデルのデータとすることを特徴とする請求項10に記載の音声認識装置。

【請求項13】 雑音データと音声データとを線形スペクトル次元で加算合成したデータを用いて雑音重畳モデルを作成する雑音重畳モデル作成手段と、  
前記ケプストラム長時間平均導出手段は前記雑音重畳モデルを用いて前記非音声部のケプストラム長時間平均を求めることを特徴とする請求項10に記載の音声認識装置。

【請求項14】 前記各処理の制御プログラムを記憶した記憶部を有し、当該制御プログラムに従って各処理を実行することを特徴とする請求項10に記載の音声認識装置。

【請求項15】 前記音声認識手段は、隠れマルコフモデルによる音声認識を行うことを特徴とする請求項11に記載の音声認識装置。

【請求項16】 前記認識結果を表示する表示手段を有することを特徴とする請求項11に記載の音声認識装置。

【請求項17】 前記認識結果を印字する印字手段を有することを特徴とする請求項11に記載の音声認識装置。

【請求項18】 前記音声認識装置は前記各処理の制御プログラムを記憶した記憶媒体とすることを特徴とする請求項10に記載の音声認識装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、隠れマルコフモデル(HMM)を用いて音声認識をおこなう、音声認識方法及び装置に関するものである。

【0002】

【従来の技術】実環境において音声認識を行う場合、特に問題となるのはマイクや電話回線特性などの影響による回線特性の歪みと、内部雑音などの加算性雑音である。これらに対処する方法として、これまでCepstrum Mean Subtraction (CMS) 法やParallel Model Combination (PMC) 法が提案されている。CMS法は「Rahim, et al.: Signal Bias Removal for Robust Telepho ne Based Speech Recognition in Adverse Environment s, Proc. of ICASSP '94, (1994).」などに詳しく、PMC法については「M. J. Gales, S. Young: An Improved Approach to theHidden Markov Model Decomposition of Speech and Noise, Proc. of ICASSP' 92, 1-233-236, (19

92) .」に詳しく述べられている。

【0003】CMS法は回線特性の歪みを補償するための一手法である。これに対しPMC法は加算性雑音に対処するための方法である。いずれも入力音声から雑音部と音声部を検出し、その情報をもとに回線歪みや雑音のない環境で作成されたHidden Markov Model (HMM) を修正し、入力環境に適應させる。これによって回線特性や雑音が変動した場合でも、柔軟に対処できる。

【0004】CMS法はインパルス応答の畳み込みで作用する乗算性雑音(回線歪み)を補償する方法である。入力音声の長時間スペクトルを入力音声から差し引き、またモデル作用に用いた音声の長時間スペクトルをモデルから差し引くことにより回線特性の差を正規化する。正規化処理は対数スペクトル領域やケプストラム領域で行うのが一般的である。乗算性雑音はこの二者の領域では加法的歪みとしてあらわれるので、引き算により雑音補償が可能となる。このうちケプストラム領域で行う方法がCMSと呼ばれている。

【0005】PMC法は無雑音環境で収録した音声で学習したHMM(音声HMM)と雑音で学習したHMM(雑音HMM)と加算合成して、モデルをより雑音重畳環境に近づける方法である。PMCにおける雑音処理では、線形スペクトル領域で雑音と音声の加算性が成立することを仮定している。一方、HMMは音声の特徴量として、ケプストラムなど対数スペクトル系のパラメータを用いることが多い。PMC法では、これらのパラメータを線形スペクトル領域に変換し、音声HMMおよび雑音HMMから得られる特徴量の線形スペクトル領域での加算合成を行っている。音声と雑音との合成後、逆変換を行って線形スペクトル領域からケプストラム領域に戻すことによって、雑音重畳音声HMMを得ている。

【0006】

【発明が解決しようとする課題】以上述べたようにCMS法を用いることにより、マイクや電話回線特性などの影響による回線特性の歪みに対処することができる。またPMC法を用いることにより、内部雑音などの加算性雑音に対処することができる。よってCMSとPMCを組み合わせることにより、回線歪みと加算性雑音の影響がある場合でも同時に適應できる。しかし、単に両者を組み合わせた場合、SNR(信号対量子化雑音比)が低い場合回線特性の推定がうまく行かない恐れがある。

【0007】これは回線特性推定の際に加算性雑音が悪影響を及ぼし、回線特性の推定がうまく行かないことに起因する。特に加算性雑音の特性が白色雑音でない場合が大きい。

【0008】

【課題を解決するための手段】上記課題を解決する為に、本発明は音声部と非音声部とを含む音声を入力し、前記入力音声に含まれる音声部から音声部のケプストラ

ム長時間平均を求め、前記入力音声に含まれる非音声部から非音声部のケプストラム長時間平均を求め、前記音声部のケプストラム長時間平均から前記非音声部のケプストラム長時間平均を線形スペクトル次元上で差し引く音声認識方法及び装置を提供する。

【0009】上記課題を解決する為に、本発明は好ましくは前記差し引いて求めたデータにより入力音声を確認する。

【0010】上記課題を解決する為に、本発明は好ましくは前記ケプストラム長時間平均を隠れマルコフモデルのデータとする。

【0011】上記課題を解決する為に、本発明は好ましくは雑音データと音声データとを線形スペクトル次元で加算合成したデータを用いて雑音重畳モデルを作成し、前記雑音重畳モデルを用いて前記非音声部のケプストラム長時間平均を求める。

【0012】上記課題を解決する為に、本発明は好ましくは前記各処理の制御プログラムを記憶した記憶部を利用し、当該制御プログラムに従って各処理を実行する。

【0013】上記課題を解決する為に、本発明は好ましくは前記音声認識は、隠れマルコフモデルによる音声認識とする。

【0014】上記課題を解決する為に、本発明は好ましくは前記認識結果を表示部に表示する。

【0015】上記課題を解決する為に、本発明は好ましくは前記認識結果を印字部に印字する。

【0016】上記課題を解決する為に、本発明は好ましくは前記各処理の制御プログラムを記憶した記憶媒体を利用し、当該制御プログラムに従って各処理を実行する。

【0017】

【発明の実施の形態】図2は本発明の音声認識装置の構成を表わすブロック図である。101はCRTや液晶表示器等の表示部であり、本発明の音声認識の結果得られる文字列を表示する。102はLBPやインクジェットプリンタ等の印字部であり、本発明の音声認識の結果得られる文字列を印字する。103はマイクロホン等の音声入力部であり、公衆回線等を介して入力するものであっても良い。104はCPUであり、記憶部105やCDROM等の着脱可能な記憶媒体107に記憶されている制御プログラムに従って以下説明する各種処理を実行するよう制御する。105は装置内に備える記憶部であって、CPU104が実行する後述の各種処理の為に制御プログラムやその為に必要な各種パラメータ(例えば音声認識に用いる辞書に相当するデータ等)、認識する音声データ等を記憶する。106は通信I/Fであって、公衆回線やLAN等の通信手段を介してデータの受授を行うべく通信の制御を行う。この通信I/Fにより、他の装置で入力された音声や、他の装置が記憶している制御プログラムや各種パラメータを本装置にとり込

10

20

30

40

50

み、記憶部105に記憶させた後、以下説明する各処理を開始するようにしても良い。107は例えばCDROM、FD等の本体に着脱可能な記憶媒体であって、記憶部105に記憶されているとして先に説明した制御プログラム、各種パラメータ、音声データを記憶することの可能な本体に着脱可能な記憶媒体であって、以下説明する各処理を開始する前にこの記憶媒体107から記憶部105にデータをダウンロードしても良いし、或いはCPU104が直接記憶媒体107にアクセスしても良い。図3は記憶部105或いは記憶媒体107に記憶されているデータのメモリマップであり、401~403に示すような各処理の制御プログラムが格納されている。また、図3には図示していないが、この他にパラメータ格納部、音声データ格納部、ワーキングエリア等も備える。

【0018】以下、図1に従って本発明の実施の形態を詳細に説明する。図1は本発明音声認識処理の機能的なブロック構成図であり、データの流れがわかるようにしてある。図1において、1は学習用音声データベース上のデータからCMを計算するCM計算部であり、CM導出プログラム401-aに従ってCMを計算する。2はその結果得て記憶部105に記憶されたCM（以後CM(2)と称する）、3は学習用音声データベース上のデータからHMMを求めるためのHMM学習部、4はその結果得て記憶部105に記憶されたHMM（以後HMM(4)と称する）、5は入力音声に含まれる音声部からCMを求めるためのCM計算部であり、CM導出プログラム401-aに従ってCMを求める。6はその結果得て記憶部105に記憶されたCM（以後CM(6)と称する）、7は入力音声に含まれる非音声部から雑音HMMを学習するための雑音HMM学習部、8はその結果得て記憶部105に記憶された雑音HMM（以後HMM(8)と称する）、9は入力音声の音声部から得られたCM(6)をケプストラム領域から線形スペクトル領域へ変換するための変換部であり、線形スペクトル領域変換プログラム401-bに従って実行される。10は雑音HMM(8)をケプストラム領域から線形スペクトル領域へ変換するための変換部であり、線形スペクトル領域変換プログラム401-cに従って実行される。11は線形スペクトル領域からケプストラム領域への変換部、12はケプストラム領域から線形ケプストラム領域への変換部、13はPMC実行部であり、PMC法データ処理プログラムに従って実行される。14はPMCで得られたモデルの線形スペクトル領域表現をケプストラム領域へ変換するための変換部、15は変換して最終的に得られたHMM（以後HMM(15)と称する）、16は15のHMM(15)を用いて音声認識を行う音声認識部である。

【0019】本認識装置は入力音声の一部を取り込み、そのデータでモデルの適応をおこなう、環境適応型とし

て動作する。まず環境適応モードにおける動作の説明を行う。最初に音声入力部103より入力した入力音声は音声部と音声が入っていない非音声部に分けられる。入力音声はケプストラムなどのパラメータに変換されているものとする。まず雑音HMM学習部7により非音声部のデータを用いて雑音用のHMMを通常のBaum-Welchアルゴリズムで学習する。これにより雑音HMM(8)ができる。また入力音声の音声部のケプストラムの長時間平均を5で計算する。それぞれをケプストラム領域から線形スペクトル領域へ変換する。変換法は前述のM. J. Gales, et. al. の文献に詳しい。この場合変換は音声部のCM(6)では平均値のみを用い、雑音HMM(8)について平均値および分散の値を用いる。入力データの音声部から計算したCM(6)の線形スペクトル領域CM(6)'での表現を以下のように表わす。

【0020】

【外1】

$$CM(6)' = \bar{x}^{11a}$$

【0021】非音声部から計算したHMM(8)の線形スペクトル領域表現の平均値HMM(8)'を以下のように表わす。

【0022】

【外2】

$$HMM(8)' = \bar{n}^{11a}$$

【0023】ここで添字11aは線形スペクトル表現を示す。次に

【0024】

【外3】

$$\bar{x}^{11a} = \bar{x}^{11a} - K_1 \bar{n}^{11a} \quad \dots (1)$$

【0025】式(1)に示す計算を行い、この結果を11の変換部で線形スペクトル表現からケプストラム表現への変換を行う。これを

【0026】

【外4】

$$\bar{x}^{cep}$$

以下のように表わす。

【0027】この式(1)に示す引き算により入力音声のCM(6)の推定誤りを低減できる。ここで $K_1$ は定数である。また引数cepはケプストラム表現を表わす。

【0028】次に学習用音声データベースの一部のデータを用いてCM学習部1でCM(2)を計算する。

【0029】

【外5】

$$CM(2) = \bar{y}^{cep}$$

【0030】また同じく学習用音声データベースの一部のデータを用いて音声認識用の音素や単語などを単位としたHMM(4)をHMM学習部3で学習する。このHMMのパラメータのうち平均値HMM(4)'を以下の

ように表わす。

【0031】 $HMM(4)' = y(t)^{c'p}$

【0032】これが適応前の雑音や回線変動に対応しないHMMとなる。このHMM(4)'を用いても音声認識は可能であるが、雑音や回線変動の影響を受けた場合認識率が低下する。次に式(2)に示す計算を行う。

【0033】

【外6】

$$y(t)^{c'p} = \bar{y}^{c'p} + \bar{x}^{c'p} \dots (2)$$

【0034】これによりHMM学習部3で求められたHMM(4)のうち平均値が変換されてHMM(4)''が求められたことになる。ここで求めたHMM(4)''は雑音と回線変動のうち回線変動のみに対処したものと

なる。  
【0035】次に12でこのHMM(4)''をケプストラム表現から線形スペクトル表現に変換する。またPMC13において10から得られた雑音HMM(8)'に定数K<sub>1</sub>をかけ、この雑音HMM(8)'と回線変動のみに対応したHMM(4)''をPMC法により混合し、HMM(13)を求める。得られたHMM(13)を14でケプストラム表現に変換することにより、雑音および回線変動に適応したHMM(14)が得られる。この得られたHMM(14)を用いて音声認識部16において一般的なHMMを用いた音声認識法により音声認識をおこなう。

【0036】上述の実施の形態とは異なる実施の形態として、以下に適応モードと認識モードを別個に行わない方法について説明する。図1の構成では、適応のための音声入力と認識のための音声入力には別個に切替えてす

＊ るようになっている。これに対し適応モードを設けない構成も可能である。図2にこの構成を示す。認識対象語彙が音声入力部103より入力されるとそのデータが音声部と非音声部に分けられ18の環境適応部に渡される。この環境適応部は図1で説明した構成5～構成14と同じものである。この環境適応部(8)により17に記憶されている適応前のHMM(HMM(4)に相当)が適応されて19の適応後のHMMが得られる(HMM(14)に相当)。このHMMを用い20の音声認識部で認識対象語彙が認識される。つまり先の例とは異なり、1つの入力音声に適応と認識の両方で用いられることになる。

【0037】

【発明の効果】従来のCMSとPMCの組合せでは加算性雑音が大きい場合、CMが正確に推定されず認識率が低下するという問題があったが、本発明によれば雑音が大きい場合でも、その雑音の影響を加味してCMが求められるため、回線変動の推定の精度が高まり、ひいては認識性能の向上につながる。

【0038】また、入力音声に適応と認識の両方に用いるようにすることにより適応と認識を同時に行うため、ユーザーに適応のための発生を強いることなく環境適応が行われる。

【図面の簡単な説明】

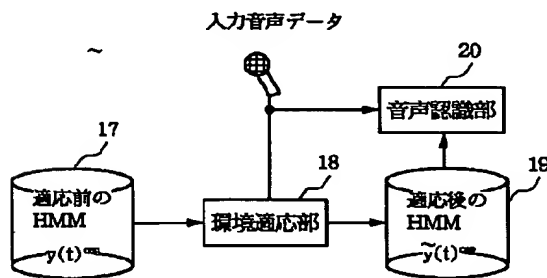
【図1】本発明の音声認識装置のブロック構成図。

【図2】他の音声認識装置のブロック構成図。

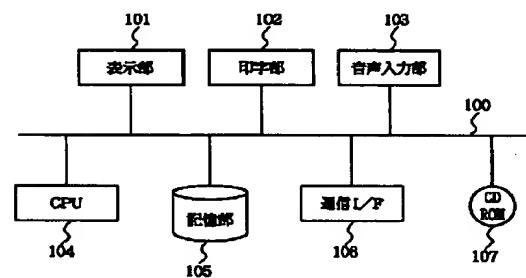
【図3】音声認識装置の構成図。

【図4】記憶部内のメモリマップ例。

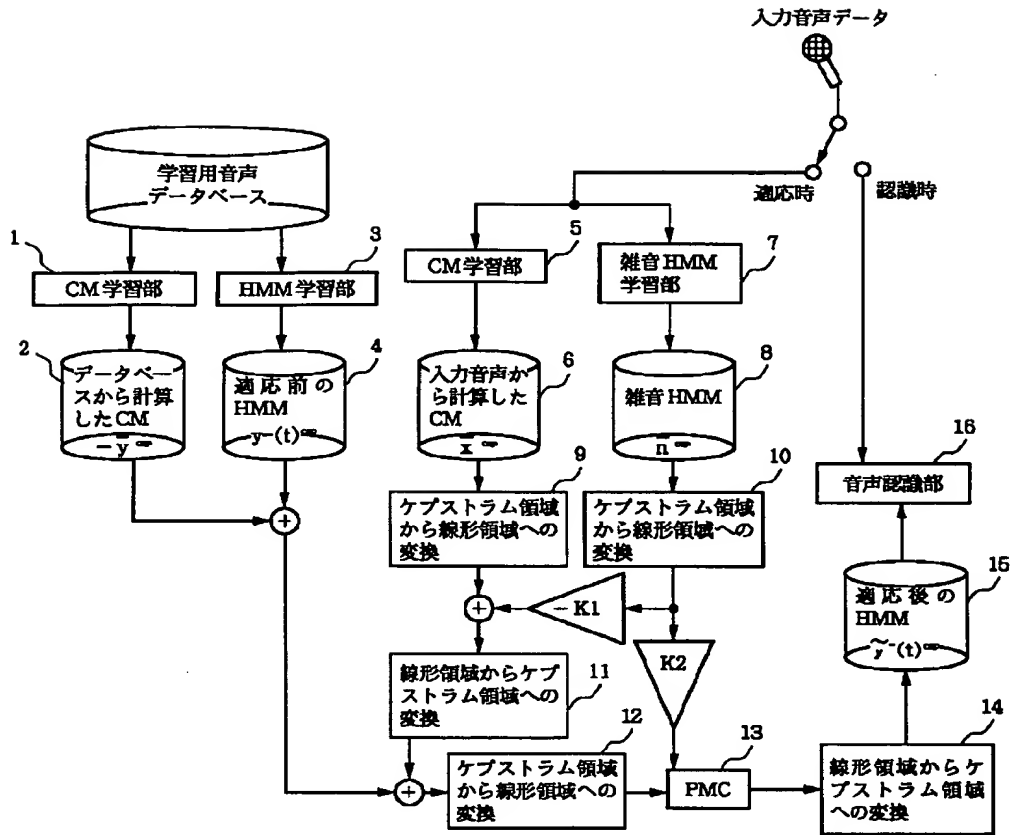
【図2】



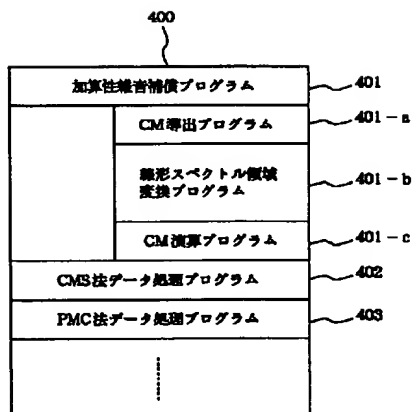
【図3】



【図1】



【図4】



【公報種別】特許法第17条の2の規定による補正の掲載  
 【部門区分】第6部門第2区分  
 【発行日】平成13年9月26日(2001.9.26)

【公開番号】特開平9-258772  
 【公開日】平成9年10月3日(1997.10.3)  
 【年通号数】公開特許公報9-2588  
 【出願番号】特願平8-68046  
 【国際特許分類第7版】

G10L 15/14  
 15/06  
 15/22  
 15/02

【F I】

G10L 3/00 535  
 521 L  
 561 C  
 9/16 301 A

【手続補正書】

【提出日】平成12年11月20日(2000.11.20)

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】0004

【補正方法】変更

【補正内容】

【0004】CMS法はインパルス応答の畳み込みで作用する乗算性雑音(回線歪み)を補償する方法である。入力音声の長時間スペクトルを入力音声から差し引き、またモデル作成に用いた音声の長時間スペクトルをモデルから差し引くことにより回線特性の差を正規化する。正規化処理は対数スペクトル領域やケプストラム領域で行うのが一般的である。乗算性雑音はこの二者の領域では加算性雑音としてあらわれるので、引き算により雑音補償が可能となる。このうちケプストラム領域で行う方法がCMSと呼ばれている。

【手続補正2】

【補正対象書類名】明細書

【補正対象項目名】0007

【補正方法】変更

【補正内容】

【0007】これは回線特性推定の際に加算性雑音が悪影響を及ぼし、回線特性の推定がうまく行かないことに起因する。特に加算性雑音の特性が白色雑音でない場合問題が大きい。

【手続補正3】

【補正対象書類名】明細書

【補正対象項目名】0017

【補正方法】変更

【補正内容】

【0017】

【発明の実施の形態】図3は本発明の音声認識装置の構成を表わすブロック図である。101はCRTや液晶表示器等の表示部であり、本発明の音声認識の結果得られる文字列を表示する。102はLBPやインクジェットプリンタ等の印字部であり、本発明の音声認識の結果得られる文字列を印字する。103はマイクロフォン等の音声入力部であり、公衆回線等を介して入力するものであっても良い。104はCPUであり、記憶部105やCDROM等の着脱可能な記憶媒体107に記憶されている制御プログラムに従って以下説明する各種処理を実行するよう制御する。105は装置内に備える記憶部であって、CPU104が実行する後述の各種処理の為に制御プログラムやその為に必要な各種パラメータ(例えば音声認識に用いる辞書に相当するデータ等)、認識する音声データ等を記憶する。106は通信I/Fであって、公衆回線やLAN等の通信手段を介してデータの受授を行うべく通信の制御を行う。この通信I/Fにより、他の装置で入力された音声や、他の装置が記憶している制御プログラムや各種パラメータを本装置にとり込み、記憶部105に記憶させた後、以下説明する各処理を開始するようにしても良い。107は例えばCDROM、FD等の本体に着脱可能な記憶媒体であって、記憶部105に記憶されているとして先に説明した制御プログラム、各種パラメータ、音声データを記憶することの可能な本体に着脱可能な記憶媒体であって、以下説明する各処理を開始する前にこの記憶媒体107から記憶部105にデータをダウンロードしても良いし、或いはCPU104が直接記憶媒体107にアクセスしても良



い。図3は記憶部105 或いは記憶媒体107に記憶されているデータのメモリマップであり、401?403に示すような各処理の制御プログラムが格納されている。また、図3には図示していないが、この他にパラメータ格納部、音声データ格納部、ワーキングエリア等も備える。

【手続補正4】

【補正対象書類名】明細書

【補正対象項目名】0018

【補正方法】変更

【補正内容】

【0018】以下、図1に従って本発明の実施の形態を詳細に説明する。図1は本発明音声認識処理の機能的なブロック構成図であり、データの流れがわかるようにしてある。図1において、1は学習用音声データベース上のデータからCM (Cepstrum Mean: ケプストラム長時間平均) を計算するCM計算部であり、CM導出プログラム401-aに従ってCMを計算する。2はその結果得て記憶部105に記憶されたCM (以後CM (2) と称する)、3は学習用音声データベース上のデータからHMMを求めるためのHMM学習部、4はその結果得て記憶部105に記憶されたHMM (以後HMM (4) と称する)、5は入力音声に含まれる音声部からCMを求めるためのCM計算部であり、CM導出プログラム401-aに従ってCMを求める。6はその結果得て記憶部105に記憶されたCM (以後CM (6) と称する)、7は入力音声に含まれる非音声部から雑音HMMを学習するための雑音HMM学習部、8はその結果得て記憶部105に記憶された雑音HMM (以後HMM (8) と称する)、9は入力音声の音声部から得られたCM (6) をケプストラム領域から線形スペクトル領域へ変換するための変換部であり、線形スペクトル領域変換プログラム401-bに従って実行される。10は雑音HMM (8) をケプストラム領域から線形スペクトル領域へ変換するための変換部であり、線形スペクトル領域変換プログラム401-cに従って実行される。11は線形スペクトル領域からケプストラム領域への変換部、12はケプストラム領域から線形ケプストラム領域への変換部、13はPMC実行部であり、PMC法データ処理プログラムに従って実行される。14はPMCで得られたモデルの線形スペクトル領域表現をケプストラム領域へ変換するための変換部、15は変換して最終的に得られたHMM (以後HMM (15) と称する)、16は15

のHMM (15) を用いて音声認識を行う音声認識部である。

【手続補正5】

【補正対象書類名】明細書

【補正対象項目名】0019

【補正方法】変更

【補正内容】

【0019】本認識装置は入力音声の一部を取り込み、そのデータでモデルの適応をおこなう、環境適応型として動作する。まず環境適応モードにおける動作の説明を行う。最初に音声入力部103より入力した入力音声は音声部と音声が入っていない非音声部に分けられる。入力音声はケプストラムなどのパラメータに変換されているものとする。まず雑音HMM学習部7により非音声部のデータを用いて雑音用のHMMを通常のBaum-Welchアルゴリズムで学習する。これにより雑音HMM (8) ができる。また入力音声の音声部のケプストラムの長時間平均を5で計算する。それぞれをケプストラム領域から線形スペクトル領域へ変換する。変換法は前述のM. J. Gales, et. al. の文献に詳しい。この場合変換は音声部のCM (6) では平均値のみを用い、雑音HMM (8) について平均値および分散の値を用いる。入力データの音声部から計算したCM (6) の線形スペクトル領域CM (6)' の表現を以下のように表わす。

【手続補正6】

【補正対象書類名】明細書

【補正対象項目名】0021

【補正方法】変更

【補正内容】

【0021】非音声部から計算したHMM (8) のパラメータのうち平均値の線形スペクトル領域表現を以下のように表わす。

【手続補正7】

【補正対象書類名】明細書

【補正対象項目名】0032

【補正方法】変更

【補正内容】

【0032】これが適応前の雑音や回線変動に対応しないHMMの平均値パラメータとなる。このHMM (4)' を用いても音声認識は可能であるが、雑音や回線変動の影響を受けた場合認識率が低下する。次に式 (2) に示す計算を行う。